数理解析研究所講究録
第 1589 巻 2008 年 229-240

An EM Algorithm for a Superposition of Markovian Arrival Processes

岡村寛之，蒲原雄也，土肥正
Hiroyuki Okamura, Yuya Kamahara and Tadashi Dohi
Department of Information Engineering, Graduate School of Engineering, Hiroshima University, Japan

1. Introduction

Since the long-range dependency of the Internet traffic was found in the mid of 1990s [16], much efforts have been spent to develop stochastic models for describing the network traffic over the classical Poisson process models. The fractional Brownian motion [21] and fractal auto-regressive integrate moving average process [11] are the typical examples which are oriented to represent the fractal nature of Internet traffic caused by the long-range dependency. As an extension of the Poisson process, Markovian arrival process (MAP) [18] and its associated stochastic process are often used to analyze mathematically the stochastic behavior arising in many practical situations such as reliability and performance evaluation.

The MAP is one of the most flexible stochastic processes, and is defined as a specific continuous-time Markov chain (CTMC). More precisely, the MAP consists of two different processes with discrete state space. One process represents the dynamics of internal state called *phase* process, another corresponds to the number of events, i.e, the counting process like a Poisson process. Here we call the number of events *level*. The phase process is usually modeled by a CTMC, and the level process is modulated by the phase process. Since the MAP is dense for any stochastic point process with an arbitrary number of phases [2], the family of MAP can be applied to approximations of complex stochastic counting processes such as the number of accesses in the Internet. In fact, Markov-modulated Poisson process (MMPP) [12], batch MMPP (BMMPP) and batch MAP (BMAP) [17], which are super- and sub-classes of MAP, have been utilized to evaluate the information communication systems based on the queueing analysis.

The MAP possesses a significant problem on the statistical inference of its parameters in practical applications. That is, we often need to determine model parameters of the MAP when evaluating the performance of real systems such as network system and production system. Given observed data in the real systems, the problem is to find appropriate parameters fitted to the observed data. The commonly used method for the parameter estimation is the maximum likelihood (ML) method. However, the ML method for MAP arises some technical difficulties due to the flexibility of MAP, i.e., a large number of free parameters are included.

To overcome this technical problem, some authors developed statistical methods to estimate the model parameters of MAP or its associated processes. The EM (expectation-maximization) algorithm [9, 28] is one of the most popular methods to estimate the parameters of MAP, and also provides a general numerical framework to derive the ML estimates for the stochastic model which involves hidden information. Since the EM algorithm has good properties on numerical computation such as a global convergence property, it is quite effective to estimate stochastic models with many free parameters; Gaussian mixed model (GMM) [5] and hidden Markov model (HMM) [4] as well as MAP.

This paper proposes an EM algorithm for a superposition of MAPs. In general, the superposition of MAPs can be formulated by a Kronecker representation of the underlying CTMC. We revisit the existing EM algorithm for an MAP [27] from the viewpoint of matrix computation, and reformulate the E-step and M-step with the Kronecker representation. This leads to reduce the matrix computation cost drastically in the EM-based parameter estimation procedure.

2. Related Work

In general, there are two approaches for fitting the family of MAP to observed data: moment-based approach and likelihood-based approach. In the moment-based approach, one determines the model parameters of MAP so as to fit theoretical moments to empirical ones from the observed data. Heffes and Lucantoni [13] provided an explicit formula for estimating the parameters of two-state MMPP by using the empirical moments of the number of arrivals. Anderson and Nielsen [1] proposed a fitting method for a superposition of 2-state MAPs based on Hurst parameter as well as the moments. Yoshihara et al. [29] developed a moment-based estimation procedure for an MMPP with several states in order to model self-similar traffic. Also, Mitchell and Liefvoort [20] developed a two-step method which deals with inter-arrival time data and lag correlation separately. The main advantage of such moment-based approaches over the likelihood-based approaches, is to reduce the computational cost.

The ML estimation for MAP has posed some difficulties until the mid of 1990s. The principle of ML estimation is to find the parameters which maximize the likelihood on the observed data as realizations of the stochastic process. The direct approach to compute ML estimates in MAPs requires large scale matrix computation. Since MAP includes numerous parameters in general, it is generally hard to find the maxima of the likelihood from the data. For example, Meier-Hellstern [19] discussed the ML estimation algorithm for a simple MMPP with only 2 phases. The EM algorithms [9, 28] for MMPP and MAP were proposed to overcome these problems.

The EM algorithm is a statistical framework to compute ML estimates under incomplete data, and is particularly useful for the stochastic models with many parameters like GMMs. The first EM algorithm for a family of MAP was the forward-backward algorithm in an HMM [4]. Deng and Mark [10] proposed a method for ML estimation of MMPP by converting an MMPP to a Markov modulated Bernoulli process (MMBP) and by applying the forward-backward algorithm in the discrete-time domain of MMBP. Asmussen et al. [3] gave an EM algorithm to estimate parameters of a phase-type (PH) distribution, and their idea could be used to estimate parameters of MMPP and MAP in the continuous-time domain. Rydén [27] extended Asmussen's idea to provide the exact ML estimates for MMPPs. In other words, the EM algorithm in Rydén [27] is analogous to the forward-backward algorithm in HMMs [4].

Based on the Rydén's work, two enhancements of EM algorithms are possible. One direction is to develop EM algorithm for a wider class of stochastic processes and data structure. Breuer [6] and Klemm et al. [15] independently discussed EM algorithms to estimate parameters of BMAPs. Okamura et al. [22, 23] developed the EM algorithm for MAP under the condition that group data of arrivals are available. Another direction is to improve the computation techniques of the original EM algorithms. Rydén's algorithm has some numerical problems on the scaling and computation of matrix exponential function. Roberts et al. [26], Klemm et al. [15] and Buchholz [7] discussed computational improvements to Rydén's algorithm. In particular, Klemm et al. [15] and Buchholz [7] implemented the uniformization technique to perform the EM algorithm effectively for the family of MAP. Furthermore, Buchholz and Panchenko [8] and Horváth et al. [14] proposed two-step fitting methods by combining the EM algorithm for PH distribution and the moment-based two-step method [20].

3. Markovian Arrival Process

3.1. Definition

The MAP is a counting process whose arrival rate is governed by a CTMC. Let $\boldsymbol{M}$ and $\boldsymbol{D}_1$ denote an infinitesimal generator of the underlying CTMC, and a rate matrix for the arrival leading to a state change of the CTMC, respectively. For instance, if there are m distinct states in the CTMC, the corresponding matrices $\boldsymbol{M}$ and $\boldsymbol{D}_1$ are given by

$$\boldsymbol{M} = \begin{pmatrix} -\mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,m} \\ \mu_{2,1} & -\mu_{2,2} & \cdots & \mu_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m,1} & \mu_{m,2} & \cdots & -\mu_{m,m} \end{pmatrix}, \quad \boldsymbol{D}_1 = \begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,m} \\ \lambda_{2,1} & \lambda_{2,2} & \cdots & \lambda_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m,1} & \lambda_{m,2} & \cdots & \lambda_{m,m} \end{pmatrix}, \tag{1}$$

where $\mu_{i,i} = \sum_{j=1,j\neq i}^{m} \mu_{i,j}$. In this paper, the state of the underlying CTMC is called a *phase*.

Let $\{N(t); t \geq 0\}$ and $\{J(t); t \geq 0\}$ be the stochastic processes which indicate the number of arrivals during time interval $[0, t)$ and the phase at time t, respectively. Define the matrix $\boldsymbol{P}_k(t)$ whose (i, j)-element is given by

$$[\boldsymbol{P}_k(t)]_{i,j} = P(N(t) = k, J(t) = j \mid N(0) = 0, J(0) = i). \tag{2}$$

Then we obtain the following differential-difference equations:

$$\frac{d}{dt}\boldsymbol{P}_0(t) = \boldsymbol{P}_0(t)\boldsymbol{D}_0, \quad \frac{d}{dt}\boldsymbol{P}_k(t) = \boldsymbol{P}_k(t)\boldsymbol{D}_0 + \boldsymbol{P}_{k-1}(t)\boldsymbol{D}_1, \quad k = 1, 2, \ldots, \tag{3}$$

where the matrix $\boldsymbol{D}_0$ is defined by

$$\boldsymbol{D}_0 = \boldsymbol{M} - \mathrm{diag}(\boldsymbol{D}_1\boldsymbol{e}). \tag{4}$$

In Eq. (4), $\boldsymbol{e}$ is a column vector whose elements are 1 and $\mathrm{diag}(\boldsymbol{D}\boldsymbol{e})$ gives a matrix with the elements of $\boldsymbol{D}\boldsymbol{e}$ on the main diagonal. Also the initial phase is determined by an initial probability vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_m)$ with $\sum_{i=1}^{m} \pi_i = 1$.

3.2. Brief overview of the EM algorithm

The EM algorithm is an iterative method for ML estimation with incomplete data [9, 28]. In general, $\mathcal{D}$ and $\mathcal{U}$ are defined as observed and unobserved data, respectively, and we estimate a set of model parameters $\boldsymbol{\theta}$ given the observed data $\mathcal{D}$. The EM algorithm is based on a procedure for finding the parameter set $\boldsymbol{\theta}$ that maximizes the expected log-likelihood function (LLF) for the complete data pair $(\mathcal{D}, \mathcal{U})$, provided that only $\mathcal{D}$ is observed. Therefore we have the following formula in the EM algorithm:

$$\boldsymbol{\theta} := \underset{\boldsymbol{\theta}}{\mathrm{argmax}}\, \mathrm{E}_{\mathcal{U}}\Big[\mathrm{LLF}(\boldsymbol{\theta}|\mathcal{D}, \mathcal{U})\Big|\mathcal{D}\Big], \tag{5}$$

where $\mathrm{E}_{\mathcal{U}}$ is the expectation operator for the unobserved data $\mathcal{U}$. In order to compute the above expected LLF, we need a provisional set of parameters. That is, Eq. (5) essentially provides an update formula of the estimated parameters, and the parameters are updated until they converge to certain values. E-step and M-step in the EM algorithm are the procedures to compute the expected LLF and to find the parameters maximizing the expected LLF, respectively.

3.3. M-step formulas for the MAP

Consider an estimation problem for an m-state MAP under a given observed data $\mathcal{D}$. The data $\mathcal{D}$ consists of K samples of time intervals between successive two arrivals, which is called the time data in this paper. More precisely, we define the time interval between the kth and the $k+1$st arrivals as t_k, and $s_0 = 0 < s_1 < \cdots < s_K$ is a cumulative time sequence, i.e., $s_k = \sum_{i=1}^{k} t_i$.

Define the following unobserved variables (random variables):

B_i: an indicator random variable for the event that the phase is i at the initial time $t = 0$.

$Y_{i,j}^{[k]}$: an indicator random variable for the event that an arrival with phase transitions from i to j occurs at time instant s_k.

$Z_i^{[k]}$: total sojourn time for phase i during a time interval (s_{k-1}, s_k).

$M_{i,j}^{[k]}$: the number of phase transitions from i to j during a time interval (s_{k-1}, s_k).

Then it is straightforward to see that

$$B_i = I(J(0) = i), \tag{6}$$

$$Y_{i,j}^{[k]} = I(J(s_k^-) = i, J(s_k^+) = j), \tag{7}$$

$$Z_i^{[k]} = \int_{s_{k-1}}^{s_k} I(J(\tau) = i) d\tau, \tag{8}$$

$$M_{i,j}^{[k]} = \int_{s_{k-1}}^{s_k} I(J(\tau^-) = i, J(\tau^+) = j) d\tau, \quad i \neq j, \tag{9}$$

where $I(\cdot)$ denotes the indicator function, and τ^- and τ^+ represent the left and right limits, i.e.,

$$I(N(\tau^-) = x, N(\tau^+) = y) = \lim_{\Delta t \to +0} I(N(\tau - \Delta t) = x, N(\tau + \Delta t) = y). \tag{10}$$

Define the parameter set $\boldsymbol{\theta} := \{\pi_i, \mu_{i,j}, \lambda_{i,j}\}$ and the unobserved variables $\mathcal{U} := \{B_i, Y_{i,j}^{[k]}, Z_i^{[k]}, M_{i,j}^{[k]}\}$ for $i, j = 1, \ldots, m$ and $k = 1, \ldots, K$. Since the parameters $\mu_{i,j}$ and $\lambda_{i,j}$ essentially equal the rates of exponential distributions representing phase transitions and arrivals in the MAP, we have the following MLEs under the complete data pair $(\mathcal{D}, \mathcal{U})$:

$$\hat{\pi}_i = B_i, \quad \hat{\mu}_{i,j} = \frac{\sum_{k=1}^{K} M_{i,j}^{[k]}}{\sum_{k=1}^{K} Z_i^{[k]}}, \quad \hat{\lambda}_{i,j} = \frac{\sum_{k=1}^{K} Y_{i,j}^{[k]}}{\sum_{k=1}^{K} Z_i^{[k]}}. \tag{11}$$

According to Eq. (5) and the above MLEs, the update formulas (M-step formulas) of the EM algorithm for MAP are obtained as follows.

$$\pi_i := \mathrm{E}[B_i|\mathcal{D}], \quad \mu_{i,j} := \frac{\sum_{k=1}^{K} \mathrm{E}[M_{i,j}^{[k]}|\mathcal{D}]}{\sum_{k=1}^{K} \mathrm{E}[Z_i^{[k]}|\mathcal{D}]}, \quad i \neq j, \quad \lambda_{i,j} := \frac{\sum_{k=1}^{K} \mathrm{E}[Y_{i,j}^{[k]}|\mathcal{D}]}{\sum_{k=1}^{K} \mathrm{E}[Z_i^{[k]}|\mathcal{D}]}, \tag{12}$$

where we omit the subscript of the expectation operation for simplicity.

3.4. E-step formulas

Define the following indicator random variables:

$$\mathcal{A}_k = I(N(s_k^+) - N(s_k^-) = 1). \tag{13}$$

Then the forward, backward and overall events can be represented by $\mathcal{F}_k = \mathcal{A}_1 \cdots \mathcal{A}_k$, $\mathcal{B}_k = \mathcal{A}_k \cdots \mathcal{A}_K$ and $\mathcal{O} = \mathcal{A}_1 \cdots \mathcal{A}_K$, respectively. For the sake of simplicity, we use the notation $P(A) = P(A = 1)$ as the probability of any indicator random variable A.

Let $\boldsymbol{f}_k(u)$ and $\boldsymbol{b}_k(u)$ be row and column vectors representing the probabilities (likelihoods) for the forward and backward events during the time period (s_{k-1}, s_k). Specifically, the i-th elements of both vectors are defined by

$$[\boldsymbol{f}_k(u)]_i = P\Big(\mathcal{F}_{k-1}, N((s_{k-1}+u)^-) - N(s_{k-1}^+) = 0, J((s_{k-1}+u)^-) = i\Big), \tag{14}$$

$$[\boldsymbol{b}_k(u)]_i = P\Big(N(s_k^-) - N((s_k-u)^+) = 0, \mathcal{A}_k, \mathcal{B}_{k+1}\Big| J((s_k-u)^+) = i\Big). \tag{15}$$

Consider the expected values in Eq. (12). By using the indicator random variable $\mathcal{O}$, we have

$$\pi_i := \frac{\mathrm{E}[B_i\mathcal{O}]}{P(\mathcal{O})} = \frac{\pi_i[\boldsymbol{b}_1(t_1)]_i}{\boldsymbol{\pi}\boldsymbol{b}_1(t_1)}. \tag{16}$$

Next we focus on computation of the expected value $\mathrm{E}[M_{i,j}^{[k]}|\mathcal{D}]$. Since $\mathrm{E}[M_{i,j}^{[k]}|\mathcal{D}] = \mathrm{E}[M_{i,j}^{[k]}\mathcal{O}]/P(\mathcal{O})$, the subsequent analysis treats only $\mathrm{E}[M_{i,j}^{[k]}\mathcal{O}]$. According to the conditional stationary independent increments of $N(t)$ provided that the Markov process $J(t)$ is known, we get

$$\begin{aligned}
\mathrm{E}[M_{i,j}^{[k]}\mathcal{O}] &= \int_{s_{k-1}}^{s_k} P\big(J(\tau^-) = i, J(\tau^+) = j, N(\tau^+) - N(\tau^-) = 0, \mathcal{O}\big)d\tau \\
&= \int_{s_{k-1}}^{s_k} P\big(\mathcal{F}_{k-1}, N(\tau^-) - N(s_{k-1}^+) = 0, J(\tau^-) = i\big) \\
&\quad \times P\big(J(\tau^+) = j, N(\tau^+) - N(\tau^-) = 0 | J(\tau^-) = i\big) \\
&\quad \times P\big(N(s_k^-) - N(\tau^+) = 0, \mathcal{A}_k, \mathcal{B}_{k+1} | J(\tau^+) = j\big)d\tau.
\end{aligned} \tag{17}$$

Using $\boldsymbol{f}_k(u)$ and $\boldsymbol{b}_k(u)$, Eq. (17) can be reduced to

$$\mathrm{E}[M_{i,j}^{[k]}\mathcal{O}] = \int_0^{t_k} [\boldsymbol{f}_k(\tau)]_i \mu_{i,j} [\boldsymbol{b}_k(t_k-\tau)]_j d\tau. \tag{18}$$

The expected value of $Z_i^{[k]}$ is obtained from Eq. (18). That is, substituting i into j in Eqs. (17) and (18) yields

$$\mathrm{E}[Z_i^{[k]}\mathcal{O}] = \int_0^{t_k} [\boldsymbol{f}_k(\tau)]_i [\boldsymbol{b}_k(t_k-\tau)]_i d\tau. \tag{19}$$

The expected value of $Y_{i,j}^{[k]}$ is also derived from the similar analysis.

$$\begin{aligned}
\mathrm{E}[Y_{i,j}^{[k]}\mathcal{O}] &= P\big(\mathcal{F}_{k-1}, N(s_k^-) - N(s_{k-1}^+) = 0, J(s_k^-) = i\big) \\
&\quad \times P\big(J(s_k^+) = j, N(s_k^+) - N(s_k^-) = 1 | J(s_k^-) = i\big) P\big(\mathcal{B}_{k+1} | J(s_k^+) = j\big).
\end{aligned} \tag{20}$$

Hence we have

$$\mathrm{E}[U_{i,j}^{[k]}\mathcal{O}] = [\boldsymbol{f}_k(t_k)]_i \lambda_{i,j} [\boldsymbol{b}_{k+1}(t_{k+1})]_j. \tag{21}$$

3.5. Computation Algorithms

To execute the EM algorithm described before, we need the vectors $\boldsymbol{f}_k(t)$ and $\boldsymbol{b}_k(t)$. In addition, the expected values $\mathrm{E}[M_{i,j}^{[k]}]$ and $\mathrm{E}[Z_i^{[k]}]$ require the computation of convolution such as Eqs. (18) and (19). In the EM algorithm for MAP, which is the so-called Rydén's method, its computation is based on the diagonalization of the matrix $\boldsymbol{D}_0$. Asmussen et al. [3] applied differential equations for solving the convolution integral for the phase-type (PH) distribution. Recently, Klemm et al. [15] and Buchholz [7] presented an improved method for the Rydén's method. The key idea is discretization of the underlying CTMC by using the uniformization technique [25]. Furthermore, Okamura et al. [24] realized the uniformization on the Asmussen's EM algorithm for PH distribution with some improvements. Here we explain the concrete E-step procedure for MAP when applying the same technique as Klemm et al. [15] and Buchholz [7].

The vectors $\boldsymbol{f}_k(t)$ and $\boldsymbol{b}_k(t)$ can be expressed as

$$\boldsymbol{f}_k(t) = \boldsymbol{\pi} \exp(\boldsymbol{D}_0 t_1)\boldsymbol{D}_1 \times \cdots \exp(\boldsymbol{D}_0 t_{k-1})\boldsymbol{D}_1 \exp(\boldsymbol{D}_0 t), \tag{22}$$

$$\boldsymbol{b}_k(t) = \exp(\boldsymbol{D}_0 t)\boldsymbol{D}_1 \times \cdots \exp(\boldsymbol{D}_0 t_K)\boldsymbol{D}_1 \boldsymbol{e}. \tag{23}$$

Therefore $\boldsymbol{f}_k(t)$ and $\boldsymbol{b}_k(t)$ can be computed by applying a simple uniformization. Let q be a constant which is larger than the maximum of absolute diagonal elements of $\boldsymbol{D}_0$. Then we have

$$\exp(\boldsymbol{D}_0 t) = \sum_{z=0}^{\infty} e^{-qt}\frac{(qt)^z}{z!}(\boldsymbol{I} + \boldsymbol{D}_0/q), \tag{24}$$

where $\boldsymbol{I}$ is the m-by-m identity matrix. The above infinite sum is truncated by a certain point in the practical computation, which is determined by the Poisson probability mass function.

On the other hand, the convolution integral is more complex for the computation based on the uniformization. Here we provide the following computation procedure as follows:

Uniformization-based Integration of Matrix Exponential:

Step 1: Compute $\boldsymbol{b}_u$ for $u = 1, \ldots, U$;

$$\boldsymbol{b}_u := \boldsymbol{P}\boldsymbol{b}_{u-1}, \quad \boldsymbol{b}_0 = \boldsymbol{b}_k(0). \tag{25}$$

Step 2: Compute $\boldsymbol{c}_u$ for $u = U - 1, \ldots, 0$;

$$\boldsymbol{c}_u := \boldsymbol{c}_{u+1}\boldsymbol{P} + e^{-qt_k}\frac{(qt_k)^{u+1}}{(u+1)!}\boldsymbol{f}_k(0), \quad \boldsymbol{c}_U := e^{-qt_k}\frac{(qt_k)^{U+1}}{(U+1)!}\boldsymbol{f}_k(0). \tag{26}$$

Step 3: Compute $\boldsymbol{H}_k = (1/q)\sum_{u=0}^{U} \boldsymbol{b}_u \boldsymbol{c}_u$,

where $q > \max_i |\mu_{i,i}|$, $\boldsymbol{P} = \boldsymbol{I} + \boldsymbol{D}_0/q$. Moreover, U is a right truncation point of uniformization satisfying

$$\sum_{u=0}^{U} e^{-qt_k}\frac{(qt_k)^u}{u!} \geq 1 - (\text{tolerance error}). \tag{27}$$

After the above computation procedure, the (j,i)-element of the matrix $\boldsymbol{H}_k$ indicates

$$[\boldsymbol{H}]_{j,i} = \int_0^{t_k} [\boldsymbol{f}_k(\tau)]_i [\boldsymbol{b}_k(t_k - \tau)]_j d\tau. \tag{28}$$

Also the time complexity of the procedure is given by $O(KUm^2)$. Compared to conventional computation methods such as diagonalization and differential equations, the time complexity can be reduced. Thus we can treat the MAP with a large number of phases by means of the above procedure.

However, in the practical situation, we encounter some difficulties to compute the expected values in the E-step, even if we use the above computation procedure. The most significant and troublesome problem is a stiffness of the underlying CTMC. In the CTMC, the stiffness corresponds to the presence of transition rates whose orders of magnitude are larger than the reciprocal of the length of the interval of integration. Intuitively, the stiff CTMC includes very rapid events and very slow events simultaneously.

The MAP is also composed of the CTMC of phase transition, and thus the stiffness problem arises in the MAP estimation. In particular when the MAP has a long-range dependency, the underlying phase process of the MAP tends to be stiff. This motivates us to develop the EM algorithm for a superposition of MAPs instead of a general MAP.

4. Superposition of MAPs

4.1. Definition

Consider a superposition of MAPs with n multiplicity. Each MAP has the set of parameters $(\boldsymbol{\pi}^{[l]}, \boldsymbol{D}_0^{[l]}, \boldsymbol{D}_1^{[l]})$ for $l = 1, \ldots, n$. Here we present the (i,j)-elements of $\boldsymbol{D}_0^{[l]}$ and $\boldsymbol{D}_1^{[l]}$ as $\mu_{i,j}^{[l]}$ and $\lambda_{i,j}^{[l]}$, respectively. In general, the superposition of MAPs can also be described by an MAP with the following parameters:

$$\boldsymbol{\pi} = \bigotimes_{l=1}^{n} \boldsymbol{\pi}^{[l]}, \quad \boldsymbol{C} = \bigoplus_{l=1}^{n} \boldsymbol{D}_0^{[l]}, \quad \boldsymbol{D} = \bigoplus_{l=1}^{n} \boldsymbol{D}_1^{[l]}, \tag{29}$$

where $\otimes$ and $\oplus$ are the Kronecker product and sum, respectively. Suppose that each MAP has m phases for the sake of simplicity. Then the number of phases of the superposition process is given by m^n. We thus represent a large size MAP by the superposition of the MAPs with a few phases. For example, it is well known that the superposition of interrupted Poisson process (IPP) can be reduced to an MMPP.

5. Parameter estimation for the superposition of MAPs

5.1. M-step formulas

Consider again the time interval data $\mathcal{D} = \{t_1, \ldots, t_K\}$, where t_k is a time interval between the $(i-1)$-st and the i-th arrivals. Also s_k is the cumulative time until the k-th arrival; $s_k = \sum_{i=1}^{k} t_i$. Yoshihara et al. [29] proposed a moment match method for a superposition of IPPs, i.e., an MMPP. In this paper, we consider the maximum likelihood estimation for the superposition of MAPs with n multiplicity, and particularly the EM algorithm is applied to the superposition of MAPs.

Similar to Section 3, we define the following unobserved values (random variables) for each superposed MAP:

$B_i^{[l]}$: an indicator random variable for the event that the phase is i at the initial time $t = 0$ in the l-th MAP.

$Y_{i,j}^{[l,k]}$: an indicator random variable for the event that an arrival with a phase transition from i to j occurs in the l-th MAP at time s_k.

$Z_i^{[l,k]}$: total sojourn time for phase i in the l-th MAP during time interval (s_{k-1}, s_k).

$M_{i,j}^{[l,k]}$: the number of phase transitions from i to j without arrivals in the l-th MAP during the time interval (s_{k-1}, s_k).

Let $J^{[l]}(t)$ and $N^{[l]}(t)$ be the phase process of the l-th MAP and the cumulative number of arrivals from the l-th MAP at time t, respectively. Then we have

$$B_i^{[l]} = I(J^{[l]}(0) = i), \tag{30}$$

$$Y_{i,j}^{[l,k]} = I(J^{[l]}(s_k^-) = i, J^{[l]}(s_k^+) = j, N^{[l]}(s_k^+) - N^{[l]}(s_k^-) = 1), \tag{31}$$

$$Z_i^{[l,k]} = \int_{s_{k-1}}^{s_k} I(J^{[l]}(\tau) = i) d\tau, \tag{32}$$

$$M_{i,j}^{[l,k]} = \int_{s_{k-1}}^{s_k} I(J^{[l]}(\tau^-) = i, J^{[l]}(\tau^+) = j) d\tau, \quad i \neq j. \tag{33}$$

Using the above unobserved variables, the estimates for the l-th MAP are given by

$$\hat{\pi}_i^{[l]} = B_i^{[l]}, \ \hat{\mu}_{i,j}^{[l]} = \frac{\sum_{k=1}^{K} M_{i,j}^{[l,k]}}{\sum_{k=1}^{K} Z_i^{[l,k]}}, \ \hat{\lambda}_{i,j}^{[l]} = \frac{\sum_{k=1}^{K} Y_{i,j}^{[l,k]}}{\sum_{k=1}^{K} Z_i^{[l,k]}}. \tag{34}$$

Therefore the M-step formulas for the superposition of MAPs are

$$\pi_i^{[l]} := \mathrm{E}[B_i^{[l]}|\mathcal{D}], \tag{35}$$

$$\mu_{i,j}^{[l]} := \frac{\sum_{k=1}^{K} \mathrm{E}[M_{i,j}^{[l,k]}|\mathcal{D}]}{\sum_{k=1}^{K} \mathrm{E}[Z_i^{[l,k]}|\mathcal{D}]}, \quad i \neq j, \tag{36}$$

$$\lambda_{i,j}^{[l]} := \frac{\sum_{k=1}^{K} \mathrm{E}[Y_{i,j}^{[l,k]}|\mathcal{D}]}{\sum_{k=1}^{K} \mathrm{E}[Z_i^{[l,k]}|\mathcal{D}]}. \tag{37}$$

The main difference between the M-step formulas in a general MAP and the superposition of MAPs is that the M-step formulas are specified for respective superposed MAPs.

5.2. E-step formulas

Define the following indicator random variables:

$$\mathcal{A}_k^{[l]} = I(N^{[l]}(s_k^+) - N^{[l]}(s_k^-) = 1), \quad \overline{\mathcal{A}_k}^{[l]} = I(N^{[l]}(s_k^+) - N^{[l]}(s_k^-) = 0). \tag{38}$$

Then one arrival event is represented by

$$\mathcal{A}_k = (\mathcal{A}_k^{[1]}\overline{\mathcal{A}}_k^{[2]} \cdots \overline{\mathcal{A}}_k^{[n]}) + \cdots + (\overline{\mathcal{A}}_k^{[1]} \cdots \overline{\mathcal{A}}_k^{[n-1]} \mathcal{A}_k^{[n]}). \tag{39}$$

Similar to the MAP case, the forward, backward and overall events can be represented by $\mathcal{F}_k = \mathcal{A}_1 \cdots \mathcal{A}_k$, $\mathcal{B}_k = \mathcal{A}_k \cdots \mathcal{A}_K$ and $\mathcal{O} = \mathcal{A}_1 \cdots \mathcal{A}_K$, respectively.

Let $\boldsymbol{f}_k(u)$ and $\boldsymbol{b}_k(u)$ be row and column vectors with m^n elements which represent the likelihoods for the forward and backward events in the time period (s_{k-1}, s_k). Thus we have

$$\boldsymbol{f}_k(u) = \bigotimes_{l=1}^{n} \boldsymbol{\pi}^{[l]} \left(\bigotimes_{l=1}^{n} \exp(\boldsymbol{D}_0^{[l]} t_1) \bigoplus_{l=1}^{n} \boldsymbol{D}_1^{[l]} \right) \cdots \left(\bigotimes_{l=1}^{n} \exp(\boldsymbol{D}_0^{[l]} t_{k-1}) \bigoplus_{l=1}^{n} \boldsymbol{D}_1^{[l]} \right) \bigotimes_{l=1}^{n} \exp(\boldsymbol{D}_0^{[l]} u) \tag{40}$$

and

$$\boldsymbol{b}_k(u) = \bigotimes_{l=1}^{n} \exp(\boldsymbol{D}_0^{[l]} u) \bigoplus_{l=1}^{n} \boldsymbol{D}_1^{[l]} \left(\bigotimes_{l=1}^{n} \exp(\boldsymbol{D}_0^{[l]} t_{k+1}) \bigoplus_{l=1}^{n} \boldsymbol{D}_1^{[l]} \right) \cdots \left(\bigotimes_{l=1}^{n} \exp(\boldsymbol{D}_0^{[l]} t_K) \bigoplus_{l=1}^{n} \boldsymbol{D}_1^{[l]} \right) \bigotimes_{l=1}^{n} \boldsymbol{e}. \tag{41}$$

Although the above equations are essentially same as those in the case of the MAP, they can reduce the time complexity of matrix operation by using the Kronecker representation, compared to the general MAP. That is, if we use Eqs. (22) and (23) for the superposition of MAP, the time complexity of the total computation of $\boldsymbol{f}_k(t_k)$, $k = 1, \ldots, K$ turns out to be $O(Km^{2n})$. In contrast, the time complexity is the Kronecker representation is given by $O(Km^2n^2)$. Accordingly, we can reduce the computation effort only by applying the Kronecker representation to the superposition of MAPs.

Next we consider the expected values in the case of the superposition of MAPs. The expected value $\mathrm{E}[B_i^{[l]}\mathcal{O}]$ is easily obtained by using $\boldsymbol{b}_k(u)$. Let $\boldsymbol{I}_i^{[l]}$ be an m^n-by-m^n matrix which consists of the Kronecker products for the identity matrices:

$$\boldsymbol{I}_{i,j}^{[l]} = \boldsymbol{I} \otimes \cdots \otimes \boldsymbol{I} \otimes \boldsymbol{e}_i \boldsymbol{e}_j^T \otimes \boldsymbol{I} \otimes \cdots \otimes \boldsymbol{I}, \tag{42}$$

where e_i is a column vector whose only i-th element is 1 (the other elements are 0), and where T is a transpose operator. Using the matrix $I_i^{[l]}$, the expected value $\mathrm{E}[B_i^{[l]}\mathcal{O}]$ is given by

$$\pi_i^{[l]} := \frac{\mathrm{E}[B_i^{[l]}\mathcal{O}]}{P(\mathcal{O})} = \frac{\bigotimes_{l=1}^{n} \pi^{[l]} I_{i,i}^{[l]} b_1(t_1)}{\bigotimes_{l=1}^{n} \pi^{(l)} b_1(t_1)}. \tag{43}$$

The expected value of $Y_{i,j}^{[l,k]}$ is similarly derived as

$$\mathrm{E}[Y_{i,j}^{[l,k]}\mathcal{O}] = \lambda_{i,j}^{[l]} f_k(t_k) I_{i,j}^{[l]} b_{k+1}(t_{k+1}). \tag{44}$$

Also, the expected value $\mathrm{E}[M_{i,j}^{[l,k]}\mathcal{O}]$ can be derived as follows.

$$\begin{aligned}
\mathrm{E}[M_{i,j}^{[l,k]}\mathcal{O}] &= \int_{s_{k-1}}^{s_k} P(J^{[l]}(\tau^-) = i, J^{[l]}(\tau^+) = j, \\
&\times N^{[1]}(\tau^+) - N^{[1]}(\tau^-) = 0, \dots, N^{[n]}(\tau^+) - N^{[n]}(\tau^-) = 0, \mathcal{O}) d\tau \\
&= \int_{s_{k-1}}^{s_k} P(\mathcal{F}_{k-1}, N^{[1]}(\tau^-) - N^{[1]}(s_{k-1}^+) = 0, \dots, N^{[n]}(\tau^-) - N^{[n]}(s_{k-1}^+) = 0, \\
&\times J^{[l]}(\tau^-) = i) P(J^{[l]}(\tau^+) = j, N^{[1]}(\tau^+) - N^{[1]}(\tau^-) = 0, \dots, \\
&\quad N^{[n]}(\tau^+) - N^{[n]}(\tau^-) = 0 | J^{[l]}(\tau^-) = i) \\
&\times P(N^{[1]}(s_k^-) - N^{[1]}(\tau^+) = 0, \dots, N^{[n]}(s_k^-) - N^{[n]}(\tau^+) = 0, \\
&\quad \mathcal{A}_k, \mathcal{B}_{k+1} | J^{[l]}(\tau^+) = j) d\tau.
\end{aligned} \tag{45}$$

The above expression seems to be quite complex, but since $N^{[l]}$, $l = 1, \dots, n$, are mutually independent, we can rewrite the equation as

$$\mathrm{E}[M_{i,j}^{[l,k]}\mathcal{O}] = \mu_{i,j}^{[l]} f_k(0) \left(\exp(D_0^{[1]} t_k) \otimes \cdots \otimes \Lambda_{i,j}^{[l,k]} \otimes \cdots \otimes \exp(D_0^{[n]} t_k)\right) b_k(0), \tag{46}$$

where $\Lambda_{i,j}^{[l,k]}$ is an m-by-m matrix;

$$\Lambda_{i,j}^{[l,k]} = \int_0^{t_k} \exp(D_0^{[l]}\tau) e_i e_j^T \exp(D_0^{[l]}(t_k - \tau)) d\tau. \tag{47}$$

The expected value of $Z_i^{[l,k]}$ is obtained as

$$\mathrm{E}[Z_i^{[l,k]}\mathcal{O}] = f_k(0) \left(\exp(D_0^{[1]} t_k) \otimes \cdots \otimes \Lambda_{i,i}^{[l,k]} \otimes \cdots \otimes \exp(D_0^{[n]} t_k)\right) b_k(0). \tag{48}$$

Based on the above equations, the computation cost of the E-step for the superposition of MAPs with n multiplicity is reduced. For instance, we can apply the computation procedure, called **Uniformization-based Integration of Matrix Exponential**, into the computation of the matrix $\Lambda_{i,j}^{[l,k]}$ directly. That is, a large number of phases are divided into the MAPs with a few phases in the computation procedure. Finally the time complexity of the E-step in the superposition of MAPs is given by $O(m^2n^2)$ even if it includes the integration of matrix exponential.

Also it should be noted that we can easily derive closed forms of the matrix exponential and its integral when the number of phases is only 2. This property may be useful to resolve the stiff Markov problem. If we represent a large-scale MAP by the superposition of 2-state MAPs, it essentially gives the closed forms of the expected values which are computed in the E-step. That is, regardless of whether the underlying phase process is stiff or not, we can compute the expected values in closed forms.

6. Numerical Example

In this section, we compare a 4-state MAP and the superposition of 2-state MAPs. In paticular we focus on the value of log likelihood function. A data set is 100 records composed of random variates given by the Weibull Distribution with shape parameter 2.0 and scale parameter 5.0. We perfome the EM algorithm for each MAP until it satisfies the termination condition. The termination condition is provided by the relative difference of log-likelihood. The algorithm stops when the relative difference between two successive log-likelihoods is lower than 1.0E-6.

The estimates in 4-state MAP are calculated as follows.

$$C = \begin{pmatrix} -2.330 & 0.8450 & 0.9540 & 0 \\ 0.9958 & -1.907 & 0 & 0.3602 \\ 0.7683 & 0 & -1.654 & 0.3357 \\ 0 & 7.966 & 11.07 & -19.04 \end{pmatrix}, \tag{49}$$

$$D = \begin{pmatrix} 0.5312 & & & \boldsymbol{O} \\ & 0.5508 & & \\ & & 0.5503 & \\ \boldsymbol{O} & & & 0 \end{pmatrix}. \tag{50}$$

Eqs. (49) and (50) show the estimation results of transition matrix and arrival matrix in 4-state MAP respectively. Moreover, Eq. (51) represents the maximum log-likelihood in 4-state MAP.

$$LLF = \ -1.62092E+02. \tag{51}$$

Next, we estimate the parameters of the superposition of 2-state MAPs. Eqs. (52) and (53) show the estiomation results of transition matrix and arrival matrix respectively. Eq. (54) represents the value of log-likelihood in MAP by the superposition of 2-state MAPs.

$$C = \begin{pmatrix} -2.0966 & 0.2694 & 1.2822 & 0 \\ 19.612 & -20.894 & 0 & 1.2822 \\ 0.1573 & 0 & -0.9718 & 0.2694 \\ 0 & 0.1574 & 19.612 & -19.770 \end{pmatrix}, \tag{52}$$

$$D = \begin{pmatrix} 0.54502 & & & \boldsymbol{O} \\ & 0 & & \\ & & 0.54503 & \\ \boldsymbol{O} & & & 0 \end{pmatrix}, \tag{53}$$

$$LLF = \ -1.62096E+02. \tag{54}$$

When the values of the log likelihood shown by Eq. (51) and (54) were compared, the value of the log likelihood was almost the same ,thus it is indicated that the superposition of 2-state MAPs is fitting data set as much as MAP in four states. Therefore, MAP in four states was equally expressible by using the superposition of 2-state MAPs.

7. Conclusions

In this paper, we have proposed an EM algorithm for the superposition of MAPs. The proposed method is essentially same as the EM algorithm for the MAP discussed in the past literature. However, in the aspect of computation cost, the proposed algorithm can reduce the time complexity, compared to the other methods. That is, by using the proposed method, we can handle the large-scale MAP in the estimation problem. In addition, if we represent the MAP with the superposition of 2-state MAPs, it also resolves the stiff problem arising in the MAP parameter estimation. In future, we will implement the proposed EM algorithm for the superposition of MAPs, and will examine the estimation performance from both accuracy and computation time viewpoints.

Acknowledgments: This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (C), Grant Nos. 18510138 (2006–2008), 19510148 (2007–2008) and Young Scientists (B), Grant No. 19700065 (2007–2008).

References

[1] A. T. Andersen and B. F. Nielsen. A Markovian approach for modeling packet traffic with long-range dependence. *IEEE Journal on Selected Areas in Communications*, 16(5):719–732, 1998.

[2] S. Asmussen and G. Koole. Marked point processes as limits of Markovian arrival streams. *Journal of Applied Probability*, 30:365–372, 1993.

[3] S. Asmussen, O. Nerman, and M. Olsson. Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23(4):419–441, 1996.

[4] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic function of Markov chains. *Annals of Mathematical Statistics*, 41(1):164–171, 1970.

[5] J. Bilmes. A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden markov Models. Technical Report ICSI-TR-97-021, University of Berkeley, 1997.

[6] L. Breuer. An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. *Annals of Operations Research*, 112:123–138, 2002.

[7] P. Buchholz. An EM-algorithm for MAP fitting from real traffic data. In P. Kemper and W. H. Sanders, editors, *Computer Performance Evaluation Modelling Techniques and Tools*, volume LNCS 2794, pages 218–236. Springer, 2003.

[8] P. Buchholz and A. Panchenko. A two-step EM-algorithm for MAP fitting. In *Proc. 19th Int'l Symp. on Computer and Information Sciences, LNCS 3280*, pages 217–227. Springer, 2004.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, B-39:1–38, 1977.

[10] L. Deng and J. Mark. Parameter estimation for Markov modulated Poisson processes via the EM algorithm with time discretization. *Telecommunication Systems*, 1:321–338, 1993.

[11] A. Erramilli, O. Narayan, and W. Willinger. Fractal queueing models. In J. H. Dshalalow, editor, *Frontiers in Queueing*, pages 245–269. CRC Press, Inc, 1997.

[12] W. Fischer and K. S. Meier-Hellstern. The Markov-modulated Poisson process cookbook. *Performance Evaluation*, 18:149–171, 1993.

[13] H. Heffes and D. M. Lucantoni. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communications*, 4(6):856–868, 1986.

[14] G. Horváth, P. Buchholz, and M. Telek. A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation. In *Proc. of the 2nd Int'l Conf. on the Quantitative Evaluation of Systems*, pages 124–133. IEEE CS Press, 2005.

[15] A. Klemm, C. Lindemann, and M. Lohmann. Traffic modeling of IP networks using the batch Markovian arrival process. *Performance Evaluation*, 54:149–173, 2003.

[16] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2(1):1–15, 2 1994.

[17] D. M. Lucantoni. New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7:1–46, 1991.

[18] D. M. Lucantoni, K. S. Meier-Hellstern, and M. F. Neuts. A single-server queue with server vacations and a class of non-renewal arrival processes. *Advances in Applied Probability*, 22:676–705, 1990.

[19] K. S. Meier-Hellstern. A fitting algorithm for Markov-modulated Poisson processes having two arrival rates. *Euro. J. of Opel. Res.*, 29:370–377, 1987.

[20] K. Mitchell and A. van de Liefvoort. Approximation models of feed-forward G/G/1/N queueing networks with correlated arrivals. *Performance Evaluation*, 52(2-4):137–152, 2003.

[21] I. Norros. On the use of fractional Brownian motion in the theory of connectionless networks. *IEEE Journal on Selected Areas in Communications*, 13:953–962, 1994.

[22] H. Okamura, T. Dohi, and K. S. Trivedi. Markovian arrival process parameter estimation with group data. in submission.

[23] H. Okamura, H. Gotoh, and T. Dohi. EM algorithm for fitting Markovian arrival processes to time-interval sampling data. In *Proceedings of International Workshop on Recent Advances in Stochastic Operations Research*, pages 203–210, 2005.

[24] H. Okamura, H. Gotoh, T. Dohi, and K. S. Trivedi. A faster EM algorithm for large-scale PH distribution. in submission.

[25] A. L. Reibman and K. S. Trivedi. Numerical transient analysis of Markov models. *Computers and Operations Research*, 15:19–36, 1988.

[26] W. J. J. Roberts, Y. Ephraim, and E. Dieguez. On Rydén's EM algorithm for estimating MMPPs. *IEEE Signal Processing Letters*, 13(6):373–376, 2006.

[27] T. Rydén. An EM algorithm for estimation in Markov-modulated Poisson processes. *Computational Statistics & Data Annalysis*, 21(4):431–447, 1996.

[28] C. F. J. Wu. On the convergence properties of the EM algorithm. *Annals of Statistics*, 11:95–103, 1983.

[29] T. Yoshihara, S. Kasahara, and Y. Takahashi. Practical time-scale fitting of self-similar traffic with Markov modulated Poisson process. *Telecommunication Systems*, 17(1-2):185–211, 2001.